# 基于结构感知深度神经网络的显著性对象检测算法 *

李　鑫 [1a]，陈雷霆 [1a, 1b, 2]，蔡洪斌 [1a, 1b]

(1. 电子科技大学 a. 计算机科学与工程学院; b. 数字媒体技术四川省重点实验室，成都 611731; 2. 电子科技大学广东电子信息工程研究院，广东 东莞 523000)

**摘　要：** 由于现有的基于深度神经网络的显著性对象检测算法忽视了对象的结构信息，使得显著性图不能完整地覆盖整个对象区域，导致检测的准确率下降。针对此问题，提出一种结构感知的深度显著性对象检测算法。算法基于一种多流结构的深度神经网络，包括特征提取网络、对象骨架检测子网络、显著性对象检测子网络和跨任务连接部件四个部分。首先，在显著性对象子网络的训练和测试阶段，通过对象骨骼检测子网络学习对象的结构信息，并利用跨任务连接部件使得显著性对象检测子网络能自动编码对象骨骼子网络学习的信息，从而感知对象的整体结构，克服对象区域检测不完整问题；其次，为了进一步提高所提方法的准确率，利用全连接条件随机场对检测结果进行进一步的优化。在三个公共数据集上的实验结果表明，该算法在检测的准确率和运行效率上均优于现有存在的基于深度学习的算法，这也说明了在深度神经网络中考虑对象结构信息的捕获是有意义的，可以有助于提高模型准确率。

**关键词：** 显著性对象检测；深度学习；显著图；卷积神经网络；对象骨架检测

**中图分类号：** TP391.41　　doi: 10.3969/j.issn.1001-3695.2018.01.0064

## Salient object detection algorithm based on structure-sensitive deep neural network

Li Xin[1a], Chen Leiting[1a, 1b, 2], Cai Hongbin[1a, 1b]

(1. *a. School of Computer Science & Engineering*, *b. Digital Media Technology Key Laboratory of Sichuan Province*, *University of Electronic Science & Technology of China*, *Chengdu* 611731, *China*; 2. *Institute of Electronic & Information Engineering in Guangdong*, *University of Electronic Science & Technology of China*, *Dongguan Guangdong* 523000, *China*)

**Abstract:** Current salient object detection algorithms based on deep neural network (DNN) are usually not able to be aware of the structure of instance, making the generated saliency maps fail to cover the entire salient object region, and thus drag down the accuracy. To solve this problem, we introduced a novel multi-stream deep neural network, in which four components were integrated in a single framework: feature extractor, object skeleton sub-network, salient object sub-network and cross-domain connections. Firstly, during the learning and testing process, the salient object detection sub-network encoded the object structure which was extracted by using object skeleton detection sub-network through the cross-domain connections, so as to make the deep model be aware of the information of object structure and overcome the problem of incomplete detection of the target area. Then, to further improve the accuracy, we proposed to use a dense conditional random field based algorithm as the refinement post-process, so as to generate a more accurate saliency map as the final results. Experimental evaluations were conducted on three widely-used benchmarks and the results show that the proposed algorithm outperforms all existing DNN-based detection algorithms in accuracy and efficiency. This also indicates that integrating object structure information into deep neural network model is meaningful, which can help to improve the overall accuracy.

**Key words:** salient object detection; deep learning; saliency map; convolutional neural network (CNN); object skeleton detection

## 0　引言

显著性对象检测的目的是在某个给定场景中自动地检测出最吸引人注意力的对象区域，从而排除不重要的信息（例如：背景杂质等）。在各种计算机视觉任务中，视觉显著性机制既能提高检测的准确率，也能提高算法的运行效率。显著性对象检测被广泛运用在人工智能相关的任务中，如稠密语义匹配错误!未找到引用源。，人体姿态识别错误!未找到引用源。，对象追踪错误!未找到引用源。等。

**收稿日期：** 2018-01-19；**修回日期：** 2018-03-15　　**基金项目：** 广东省应用型科技研发专项资金资助项目（2015B010131002）；广东省科技计划资助项目（2016A040403004）；东莞市重大科技项目（2015215102）

**作者简介：** 李鑫（1986-），男，四川江油人，博士研究生，主要研究方向为计算机视觉（xinli_uestc@hotmail.com）；陈雷霆（1966-），男，重庆人，教授，博士，主要研究方向为多媒体应用；蔡洪斌（1966-），男，四川广安人，教授，博士，主要研究方向为三维图形.

显著性对象检测分为两个主要的研究方向：自顶向下（top-down）的显著性对象检测和自底向上（bottom-up）的显著性对象检测。自顶向下的显著性对象检测与对象检测相关，其目标是自动定位指定类别的对象所在位置。而自底向上的显著性对象检测算法则是自动地识别给定场景中最为显著的一个或者多个对象区域。相比目标驱动的、自顶向下的显著性对象检测算法，自底向上的方法具有下意识、快速以及数据驱动的特征。本文研究自底向上的显著性对象检测算法，即自适应地检测任意场景中最为显著的对象区域。

自底向上的显著性对象检测可以主要归纳为两类：传统检测模型和基于深度学习的检测模型。传统模型错误!未找到引用源。~错误!未找到引用源。利用低级的图像特征（如颜色错误!未找到引用源。、梯度错误!未找到引用源。、深度错误!未找到引用源。等）计算显著性值，并基于简单的先验知识（如中心先验，边缘先验等）和观察进行建模[9]，很难用于较复杂的场景。

近几年，深度学习广泛地运用于计算机视觉的不同任务中。显著性对象检测受益于这种基于深度神经网络、端到端自动学习的高级语义特征。相比传统模型，基于深度模型的显著性检测算法可以利用语义信息极大提升显著性对象检测的准确率。这些基于深度学习的方法可进一步分为基于区域的深度学习检测和基于像素级的深度学习检测。基于区域的显著性检测算法将每个区域当成一个独立的单元进行计算。Li 等人错误!未找到引用源。提出一个多尺度显著性对象检测网络去计算每个超像素的显著性值。Wang 等人错误!未找到引用源。把从不同网络提取出来的局部和全局显著性特征融合到统一框架中，然后计算每个区域的显著性值。类似地, Zhao 等人错误!未找到引用源。利用深度神经网络同时提取局部和全局的上下文信息，然后计算每个区域的显著性值。然而，目前的基于区域的深度检测模型都将每个区域看成一个独立的单元，造成计算上的冗余，降低了运行效率。针对这个问题，基于像素级的显著性对象检测通过深度神经网络模型，直接求得每个像素的显著性值。Wang 等人错误!未找到引用源。把先验知识用于循环网络，然后直接预测每个像素的显著性值。Hou 等人错误!未找到引用源。提出同时融合多个尺度信息的全卷积网络去检测显著性对象。这个方法基于嵌套网络框架，通过内部的连接去整合多尺度信息。Li 等人错误!未找到引用源。提出用级联结构网络同时捕获多尺度上下文信息，以解决对象的尺度问题。

但由于对象的多样性、训练数据库的样本偏差等，在显著性对象结构复杂的情况下，现有的深度模型，很难检测出完整的对象区域。如何在网络学习和测试的过程中意识和捕获到对象的完整结构信息，是被过去和现有研究所忽略的问题。针对这些问题，本文提出一种基于结构感知的深度神经网络模型，并将之运用于检测场景中的显著性对象。

## 1 本文方法

为了在训练和检测中能使得深度神经网络自动捕获和感知对象的整体结构信息，然后利用这些结构信息完整地检测出整个显著性对象区域，本文提出一种基于结构感知的深度神经网络模型。如图 1 所示，该神经网络模型包括四个部分：特征提取网络，对象骨骼提取子网络，显著性对象检测子网络和跨任务连接部件。本文提出的深度模型贡献在于：a)不同于传统多任务网络错误!未找到引用源。，仅仅通过共享底层表达的方法互相传递信息，本文的网络把跨任务连接部件用在多个尺度上，并深度融合对象骨骼子网络学习的知识和信息，使得显著性检测子网络在特征学习和编码的过程中能主动吸收、整合不同尺度的对象骨骼信息，以便更好地捕捉对象结构特征；b)通过共享的底层表达，对象骨骼检测子网络能共享对象显著性信息，使对象骨骼的检测集中在显著的对象上，从而克服背景杂质的干扰；c)利用全连接条件随机场（dense conditional random field, DenseCRF）对深度网络的检测结果进行进一步的优化，使得检测结果更准确。

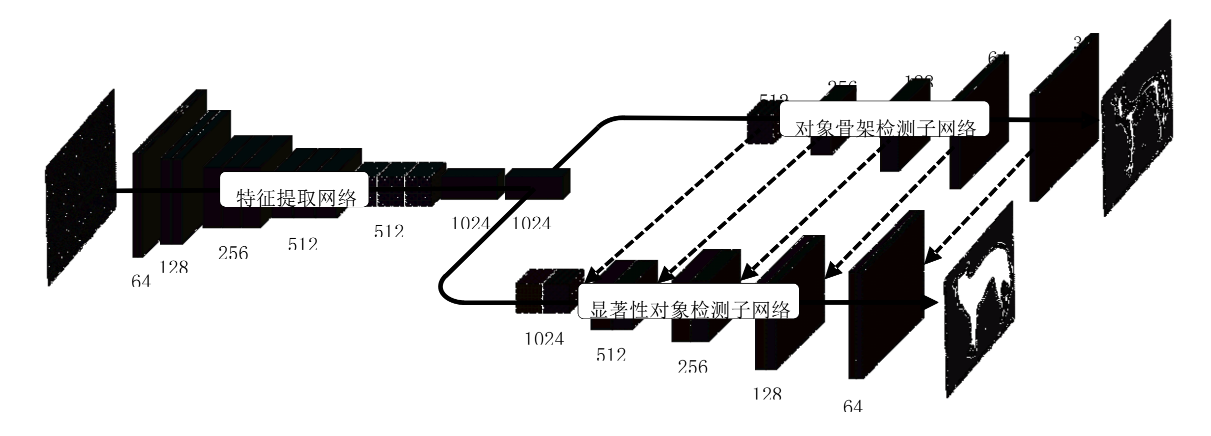



**图 1 基于结构感知的深度神经网络的整体结构**

特征提取网络基于 VGG-16 网络结构错误!未找到引用源。。传统的 VGG-16 网络结构包含 5 组不同尺度的卷积层以及 3 个一维向量。由于本文的目标是像素级的检测，因而设计的特征提取网络只取 VGG-16 的前五组卷积层。另外，为了确保网络深度，模型添加了一个额外的卷积层，维度设为 1024 维。除此之外，一个额外的反卷积层也被添加在特征提取网络中，用作后续子网络（包括显著性对象检测子网络和对象骨骼检测子网络）的共享底层表达。特征提取网络的前五组卷积层的参数初始化利

用在 ImageNet 中训练的 VGG-16 对应的网络参数，而新添加的两个卷积层的参数被随机初始化。特征提取网络的输入为任意一张 RGB 图 $I$，输出为提取的深度特征，标记为 $F$，作为后续子网络的底层表达。

### 1.1　对象骨骼检测子网络

对象骨骼检测子网络的输入为特征提取网络所提取的深度特征图 $F$，输出为对应的对象骨骼图 $S_k(F;\theta_{sk})$，其中 $\theta_{sk}$ 为对象骨骼提取子网络的网络参数。对象骨骼提取网络基于反卷积网络结构，包括 5 个不同尺度的逆池化层（up-pooling layer）和反卷积层（De-Convolution layer, Deconv），每组的尺度与特征提取网络的尺度对应，从而最

终获取的对象骨骼图的尺度和输入尺度大小一致。如图 2 所示，不同于池化层用于整合信息、缩小感受野、排除杂质，逆池化层具有相反的操作，通过重新定位池化区域最大值的位置重建被池化信息。其中逆池化操作和反卷积操作与池化和卷积操作具有完全相反的步骤，用于重建信息。

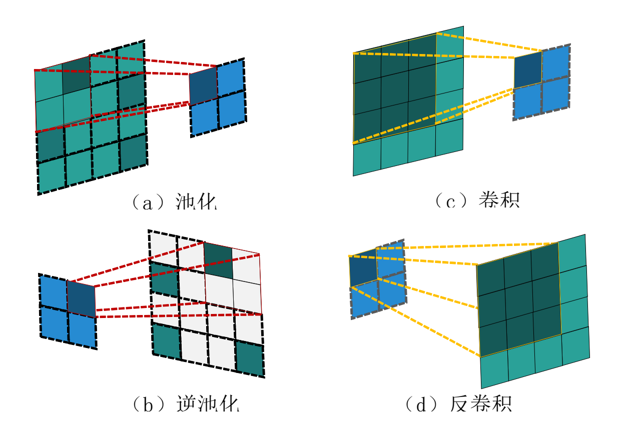



**图 2　逆池化操作和反卷积操作示意图特征提取网络**

由于逆池化操作的输出为一个扩大而稀疏的特征，反卷积操作通过类似于卷积的操作去进一步扩充丢失的信息。另外，每个反卷积层后，额外添加了一个 Relu 激活函数（Relu activation function）。为了防止过度拟合，对每个 Relu 层后添加一个 Dropout 层。对象骨骼检测子网络的配置如表 1 所示。

表 1　对象骨骼检测子网络配置

| 反卷积层 | 设置 | 核参数设置 | 激活函数 |
|---|---|---|---|
| DeConv1 | DeConv&Unpool | 5×5×512 | Relu |
| DeConv2 | DeConv&Unpool | 5×5×256 | Relu |
| DeConv3 | DeConv&Unpool | 5×5×128 | Relu |
| DeConv4 | DeConv&Unpool | 5×5×64 | Relu |
| DeConv5 | DeConv&Unpool | 5×5×32 | Relu |

对象骨骼检测子网络的任务是提取目标图像中对象的骨骼信息。网络参数 $\theta_{sk}$ 的训练属于像素级的回归问题（pre-pixel regression）。基于此，本文构造了如下目标函数：

$$L_S(\theta_{sk})=\min_{\theta_s}\sum_j \mathrm{l}_{sk}\{S_{k-gt}{}^j(p),S_k{}^j(F;\theta_{sk})(p)\} \quad (1)$$

其中：$S_{k-gt}{}^j(p)$ 表示在第 $j$ 个训练样本，像素坐标 $p$ 处的用户标注（Ground Truth, GT）结果；$S_k{}^j(F;\theta_{sk})(p)$ 表示深度模型根据网络参数 $\theta_{sk}$ 预测的第 $j$ 个训练样本，坐标 $p$ 处的值。$\mathrm{l}_{sk}\{\bullet\}$ 表示损失函数，用于计算深度模型预测结果和用户标注结果的差异。这里，损失函数为典型的交叉熵代价函数，定义如下：

$$\begin{aligned}\mathrm{l}_{sk}&=-\sum_{p\in Z}S_{k-gt}(p)\log P(S_k(p)=1\,|\,F;\theta_{sk})\\&+(1-S_{k-gt}(p))\log P(S_k(p)=0\,|\,F;\theta_{sk})\end{aligned} \quad (2)$$

其中：$S_{k-gt}(p)$ 表示在坐标 $p$ 处的用户标注，$S_k(p)$ 表示用户预测的坐标 $p$ 处的结果。

对象骨骼检测子网络通过大量样本的训练学习到反映对象整体结构的骨骼信息。本文的目标是利用该信息进一步提高显著性对象检测的完整性和准确率，下一节将介绍如何运用学习的对象骨骼信息去帮助学习显著性对象信息。

### 1.2　显著性对象检测子网络

显著性对象检测子网络和对象骨骼检测子网络都基于相同的底层表达 $F$，输出为一个显著性图 $S_{sal}(F;\theta_{sal})$，其中 $\theta_{sal}$ 表示对象检测子网络的网络参数。显著性对象子网络的网络参数配置如表 2 所示。

表 2　显著性对象检测子网络配置

| 反卷积层 | 设置 | 核参数设置 | 激活函数 |
|---|---|---|---|
| DeConv1 | DeConv&Unpool | 3×3×512 | Relu |
| DeConv2 | DeConv&Unpool | 3×3×256 | Relu |
| DeConv3 | DeConv&Unpool | 3×3×128 | Relu |
| DeConv4 | DeConv&Unpool | 3×3×64 | Relu |
| DeConv5 | DeConv&Unpool | 3×3×32 | Relu |

不同于对象骨骼检测子网络结构，显著性对象检测子网络利用跨任务连接部件同时编码显著性对象信息和对象骨骼信息，帮助学习对象结构感知的显著性特征。具体来说，显著性检测子网络在多尺度的学习过程中与相应尺度的对象骨骼学习得到的特征进行整合：

$$F_{s_i}^{sal}=\sigma\{cat(f_{s_i}^{sal},\ f_{s_i}^{sk})\otimes W_{s_i}+b_{s_i}\} \quad (3)$$

其中：$\otimes$ 表示卷积操作，$W_{s_i}$ 表示尺度 $S_i$ 的卷积（convolutional filters），$b_{s_i}$ 表示尺度 $S_i$ 的偏置（biases）。本文运用 Relu 作为非线性激活函数 $\sigma\{\bullet\}$。$cat(\bullet)$ 表示一个连接函数，用来连接两个不同的特征图（feature maps）。如图 1 所示，两个不同子网络学习的特征通过跨任务连接部件（cross-domain connections）整合为一个信息更加丰富的特征图，显著性对象检测子网络更高尺度的特征学习基于整合后的特征，从而能编码对象骨骼检测子网络学习的知识。值得注意的是：上述两种信息的整合为非线性的，由深度神经网络通过端到端地训练自动获取。

类似于对象骨骼检测子网络，显著性对象检测子网络的参数训练也是一个向用户标注数据的回归问题。目标函数构造为：

$$L_{SAL}(\theta_{sal})=\min_{\theta_{sal}}\sum_{j}\mathrm{l}_{sal}\{S_{sal-gt}{}^{j}(q),S_{sal}{}^{j}(F;\theta_{sal})(q)\} \quad (4)$$

其中：$S_{sal-gt}{}^{j}(q)$ 为第 $j$ 个训练样本，像素坐标 $q$ 处的用户标注的显著性值。$S_{sal}{}^{j}(F;\theta_{sal})(q)$ 表示显著性对象检测子网络预测的第 $j$ 个训练样本，坐标 $q$ 处的显著性值。损失函数也选用交叉熵代价函数：

$$\begin{aligned}\mathrm{l}_{sal}=&-\sum_{p\in Z}S_{sal-gt}(q)\log P(S_{sal}(q)=1\,|\,F;\theta_{sal})\\&+(1-S_{sal-gt}(q))\log P(S_{sal}(q)=0\,|\,F;\theta_{sal})\end{aligned} \quad (5)$$

其中：$S_{sal-gt}(q)$ 表示图像坐标 $q$ 处用户标注的显著性值，$S_{sal}(q)$ 表示显著性对象子网络的预测值。

骨骼信息反映了对象的整体结构信息，显著性对象检测子网络在特征学习的过程中，以端到端的方式学习如何整合对象骨骼信息，从而感知对象结构，提高检测结果的完整性。相比单任务反卷积网络，本文在深度神经网络中加入骨骼信息的学习，且将多个尺度的信息融合，从而提高了检测结果的完整性。如图 3 所示，对于结构复杂的对象，本文模型依然可以准确检测出完整的显著性对象区域。而单任务的反卷积网络，由于不能感知对象结构，则不能很好地保存检测对象区域的完整性。其中 **GT 表示用户标注。**

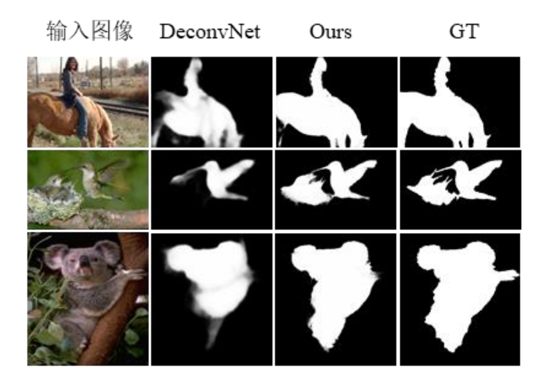


**图 3 单任务反卷积网络（DeconvNet），本文模型结果对比**

由于目前没有可以同时包含对象骨骼和显著性对象区域信息的训练数据集，因此不能够直接训练本文提出的网络模型。为了解决这个问题，本文利用一种交替学习的方式完成网络的训练。主要思想是运用不同的训练数据集交替地训练显著性检测神经网络的两个不同子网络参数，直至目标函数式（1）（4）均收敛。

## 1.3 显著性图优化方法

尽管提出的深度模型已经可以获得准确的检测结果，但是一些极端情形下，深度神经网络模型仍然不能很好地保存对象的边缘信息。为了克服这个问题，本文利用全连接条件随机场（Dense CRF）进一步优化深度模型的检测结果 $S_{sal}$。类似于语义分割，定义如下能量方程：

$$E(L)=-\sum_{i}\log P(l_i)+\sum_{i,j}\omega_{i,j}(l_i,l_j) \quad (6)$$

上述能量方程由数据项和平滑项组成。其中 $L$ 表示显著性值的标签。$L\in\{0,1\}$，当 $L(i)=1$ 时表示像素 $i$ 属于显著性对象区域，反之当 $L(i)=0$ 表示该对应像素属于非显著性对象区域。$\log P(l_i)$ 为数据项，其中 $P(l_i)$ 的值为深度神经网络预测像素 $i$ 的显著性值，因而 $P(1)=S_{sal}(i)$ 且 $P(0)=1-S_{sal}(i)$。$\omega_{i,j}(l_i,l_j)$ 为平滑项，用于鼓励颜色相近的相邻像素具有相同的标签。平滑项的定义如下：

$$\begin{aligned}\omega_{i,j}(l_i,l_j)=\mu(l_i,l_j)\Bigg\{&\gamma_1\exp\left(-\frac{\|p_i-p_j\|^2}{2v^2}-\frac{\|I_i-I_j\|^2}{2o^2}\right)\\&+\gamma_2\exp\left(-\frac{\|p_i-p_j\|^2}{2\kappa^2}\right)\Bigg\}\end{aligned} \quad (7)$$

其中：$p_i$ 和 $p_j$ 用于表示像素 $i$ 和像素 $j$ 所对应的坐标位置，$I_i$ 与 $I_j$ 分别表示像素 $i$ 和像素 $j$ 所对应的像素亮度。另外，当 $l_i\neq l_j$ 时，$\mu(l_i,l_j)=1$；反之，$\mu(l_i,l_j)=0$。$v$、$o$ 以及 $\kappa$ 为能量方程的权重，用于分别控制位置和颜色的相似性等对优化方程的影响。

求解式（7）为一个 NP-hard 问题，这里简单采用平均差

近似方法求解。优化方程将原本模型输出的显著性图作为输入，并且进行优化，得到一个更加准确的结果作为输出。该优化方法能更很的保存对象的边缘信息以及对象区域连续性。

## 2 实验和分析

### 2.1 实验环境

本实验的训练和测试环境为：Ubuntu 14.04， Caffe 框架，并且利用 NVIDIA GTX Titan X GPU 加速深度神经网络。深度神经网络训练的数据集为：MSRA-10[10] 作为显著性对象检测子网络的训练集 $\{I_{Sal}, G_{Sal}\}$；Pascal-SYM[21] 作为对象骨骼检测子网络的训练集 $\{I_{Ske}, G_{Ske}\}$。

测试集采用了三个公共数据集，包括 ECSSD 错误!未找到引用源。、DUT-OMRON 错误!未找到引用源。以及 HKU-IS 错误!未找到引用源。。其中，ECSSD 为目前最为广泛使用的公共数据集。该数据集包括 1,000 张具有复杂背景的图像。DUT-OMRON 为一个大型公共数据库，包含 5,168 张非常挑战的图像。这些图像具有一个或者多个显著性对象，这些显著性对象具有各异的姿态和外形。该数据库也被认为是目前最具难度的数据库。HKU-IS 为近期发布的公共数据集。该数据集包含 1,447 张图像。该数据构建时候特意避免了显著性对象一些简单的先验知识，比如对比度、中心先验等，并且该数据集中的显著性对象一般为多个。因而该数据库也是目前非常有难度的数据库。上述数据库可以用于验证显著性模型的准确性和鲁棒性。

### 2.2 训练和测试

训练和测试的步骤具体如下：

**a)**在训练和测试过程中，先将图像尺度统一调整为 224×224，以保证 VGG-16 网络的输入要求和运行效率；

**b)**在 ImageNet 上训练 VGG-16 网络，将得到的网络参数去初始化本文模型中的特征提取网络的前五组卷积层参数，而新添加的卷积层以及反卷积层的网络参数以正态分布来随机初始化。

**c)**通过两个数据集：显著性对象检测数据集 $\{I_{Sal}, G_{Sal}\}$ 和对象骨骼检测子网络的训练集 $\{I_{Ske}, G_{Ske}\}$ 分别训练两个不同的子网络；

**(a)**固定特征提取网络前五组卷积层的网络参数，以及显著性对象检测子网路参数 $\theta_{sal}$，数据集 $\{I_{Ske}, G_{Ske}\}$ 首先被用来训练对象骨骼检测子网络的参数 $\theta_{sk}$；

**(b)** 固定对象骨骼检测子网络参数 $\theta_{sk}$，利用数据集 $\{I_{Sal}, G_{Sal}\}$ 训练显著性对象检测子网络的参数 $\theta_{sal}$；

(c) 上述步骤 (a)(b) 均采用 “poly” 衰减算法，根据 $(1-{}^{iter}/_{max_{iter}})^{power}$ 自动调整学习率。其中 power 被设置为 $10^{-7}$，并且最大训练次数 $\max_{iter}$ 设置为 60,000；

**(d)**上述步骤交替进行，利用随机梯度下降方法训练参数，直至目标函数收敛。

d)测试的过程中，将所需要测试的图像作为输入放入训练好的神经网络中，输出初始显著性图。

e)深度模型输出的初始显著性图运用全连接条件随机场，即式（6），进一步优化得出最终检测的显著性图。

### 2.3 验证方法

为了验证本文所提方法的有效性和准确性，本文采用四种标准验证方法进行准确性和鲁棒性验证，包括：精确率-召回曲线（PR-curves）、F-measure、weighted F-measure 以及绝对方差错误（MAE）。

首先，可以根据生成的显著性图，基于不同的阈值将其转换为一系列二值图 $B$，并且比较它们和用户标注 $G$ 的差异。因而可以获得给定显著性图的准确率。$P=\frac{|B\cap G|}{|B|}$，以及召回率 $R=\frac{|B\cap G|}{|G|}$。之后，根据不同阈值计算精确率-召回曲线。F-measure（$F_\beta$）的计算如下：

$$F_\beta=(1+\beta^2)\frac{P\times R}{\beta^2 P+R} \quad (8)$$

根据相关文献**错误!未找到引用源。,错误!未找到引用源。**的建议，上式中，$\beta^2$ 设置为 0.3。F-measure 指标反映了方法的准确性，除此之外，weighted F-measure（$F_\beta^\omega$）也被用于验证显著性检测算法的准确性，计算如下：

$$F_\beta^\omega=(1+\beta^2)\frac{P^\omega\times R^\omega}{\beta^2 P^\omega+R^\omega} \quad (9)$$

其中，$P^\omega$ 和 $R^\omega$ 为加权准确率（weighted precision）和加权召回率（weighted recall）。

绝对方差错误则用于计算每个像素和用户标注（ground truth）的平均误差，计算如下：

$$MAE=\frac{1}{W\times H}\sum_{x=1}^{W}\sum_{y=1}^{H}|S(x,y)-G(x,y)|, \quad (10)$$

其中 $W$ 和 $H$ 分别代表图像的长度和宽度。

### 2.4 对比分析

为了验证本文提出算法的有效性和鲁棒性，与 5 个显著性对象检测算法进行对比，其中包括 2 个最具代表性的传统显著

性对象检测算法：基于鲁棒背景检测的显著性对象检测算法(robust background detection, RBD)[错误!未找到引用源。]，显著性区域特征融合的显著性对象检测算法(Discriminative Regional Feature Integration, DRFI) [错误!未找到引用源。]；以及 3 个基于深度学习的显著性对象检测方法，包括：基于距离图的深度显著性对象检测算法(encoded low-level distance, ELD) [错误!未找到引用源。]，基于深度全卷积网络的显著性对象检测算法(deep saliency, DS)[错误!未找到引用源。]， 基于深度对比度的显著性对象检测算法(deep contrast learning, DCL) [错误!未找到引用源。]。

2.4.1 定量分析

为了验证本文所提算法的准确性和鲁棒性，首先利用 F-measure 值比较本文所提算法和其他显著性对象检测算法的准确性。如表 3 所示，本文所提算法在三个公共数据集均获得最高的 F-measure 值。具体来说，本文所提方法在 ECSSD 数据库的结果表明本文算法可以在复杂的场景中准确检测出结构复杂的显著性对象。在 DUT-OMRON 中，本文所提算法也高于所有基于深度学习的显著性对象检测算法，F-measure 值达到了 0.765，远高于 ELD 的 0.720。该数据库为当前最大数据库，包含了各种不同的场景和对象，因此可以证明本文方法具有一定的鲁棒性，可以用于检测各种复杂场景中的显著性对象。在 HKU-IS 数据集中，本文算法也获得最高的 F-measure 值。这证明本文方法可以比当前算法更鲁邦地检测出同一场景中的多个显著性对象。根据 F-measure 的检测结果，还可以得出如下结论：a）利用深度神经网络提取的语义信息可以极大地提升显著性对象检测算法的准确性；b）本文算法由于整合了对象的结构信息，因而可以更好的检测出复杂场景中的显著性对象，使得方法更具准确性和鲁棒性；c）运用全连接条件随机场优化深度模型输出的显著性图可以进一步提高检测的准确率。

如表 3 所示，本文算法在三个公共数据集上都获得了更低的 MAE 值。具体来说，与对比算法比较，在 ECSSD、DUT-OMRON 以及 HKU-IS，本文所提方法 MAE 值分别降低了 19.4%、26.3%以及 27.1%。该验证结果说明本文算法可以生成和用户标注更为接近的结果，证明了该方法的准确性和鲁棒性。主要原因在于：a）整合对象骨骼信息可以使得深度网络感知对象结构，从而能完整的检测出整个显著性对象区域；b）运用全连接条件随机场对显著性图进一步优化可以更好的保存对象的边缘信息，使得结果更加准确。

最后，根据精确率-召回曲线（PR-curves）和 weighted F-measure（图 4），本文算法在三个公共数据集上也有更优结果。再次证明了本文所提算法的准确性和鲁棒性。

表 3　不同方法在三个公共数据集的结果比较结果。其中，"Ours"表示本文深度神经网络模型，而"Ours+"表示运用额外的 DenseCRF 进一步优化模型输出的结果。

| 方法 | ECSSD | | DUT-OMRON | | HKU-IS | |
|---|---|---|---|---|---|---|
| | $F_\beta$ | MAE | $F_\beta$ | MAE | $F_\beta$ | MAE |
| DRFI | 0.786 | 0.164 | 0.665 | 0.155 | 0.684 | 0.226 |
| RBD | 0.716 | 0.171 | 0.630 | 0.144 | 0.657 | 0.120 |
| DS | 0.882 | 0.122 | 0.716 | 0.120 | 0.866 | 0.079 |
| ELD | 0.869 | 0.098 | 0.720 | 0.091 | 0.777 | 0.121 |
| DCL | 0.887 | 0.072 | 0.718 | 0.094 | 0.879 | 0.059 |
| Ours | 0.900 | 0.068 | 0.753 | 0.077 | 0.893 | 0.056 |
| Ours+ | 0.907 | 0.058 | 0.765 | 0.067 | 0.902 | 0.043 |

2.4.2 定性分析

如图 5 所示，本文算法在检测复杂场景中的显著性对象区域时，能够更好地排除背景杂质、完整提取显著性对象区域。对比现存显著性对象方法，本文方法在应对这些相对少见的场景时，可以更准确地发现显著性对象区域，从而证明该方法具有更好的鲁棒性。也说明了，在深度神经网络中考虑对象骨骼信息的整合有利于深度神经网络更好的感知对象的形状信息，从而帮助克服二义性问题。

另外，从定性比较中，还可以得出如下结论：a）相比基于人工设计特征的显著性对象检测方法，基于深度神经网络的显著性对象检测方法可以更准确地检测复杂场景中的显著性对象；b）在深度网络模型中考虑对象的结构信息，可以帮助克服复杂场景检测时的二义性问题，提高检测方法的鲁棒性。

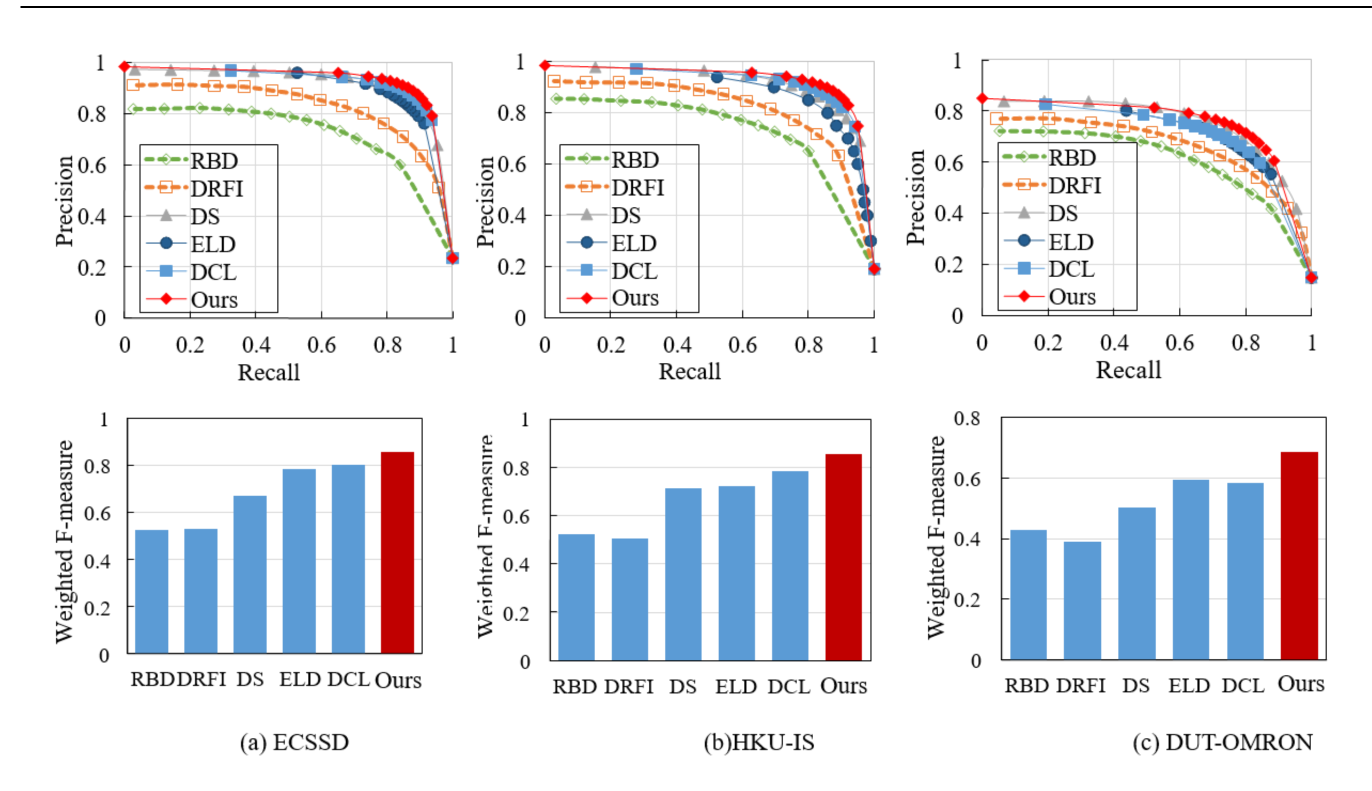



图 4　准确率-召回曲线和 weighted F-measure

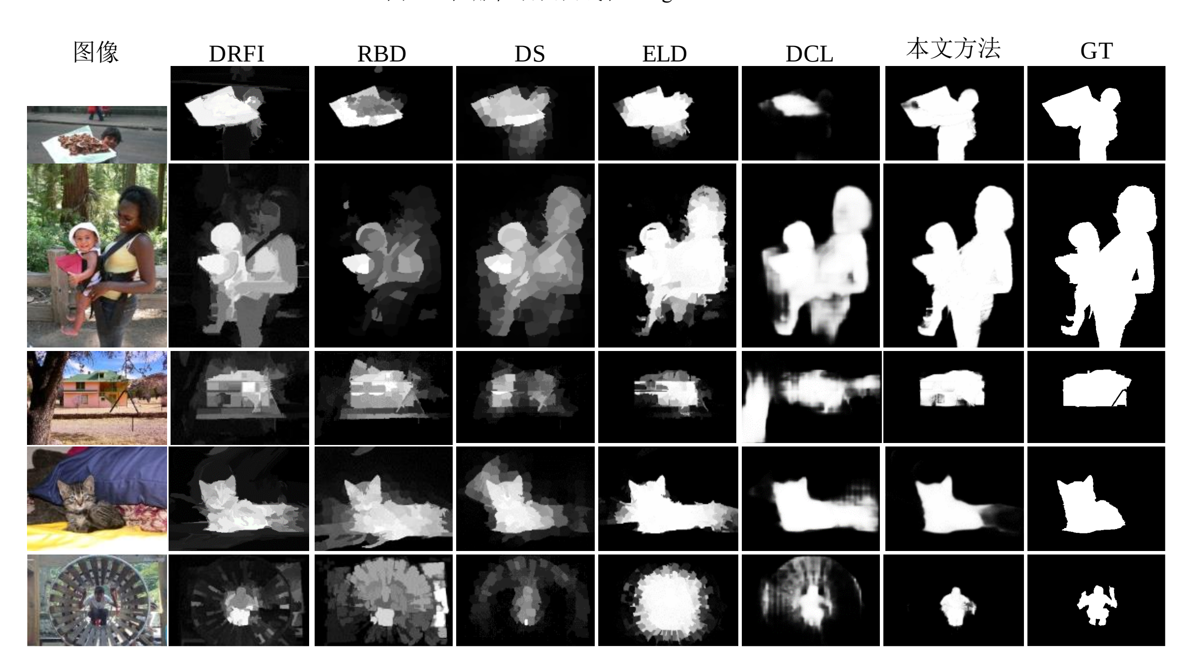



**图 5 不同方法定性比较，GT 表示用户标注结果。**

2.4.3 运行效率分析

运行效率测试结果如表 4 所示。传统方法的测试平台为 i7 2.50 GHz CPU 以及 8 GB RAM；而基于深度学习的显著性对象检测方法的测试平台为 NVIDIA GTX Titan X GPU 以及 12G 显存。实验结果表明，传统方法的运行效率低于基于深度学习的显著性对象检测算法。本文的深度模型运行效率为 30fps。但是由于本文算法的结果需要额外的后处理，因此处理一张输入图像的整体运行时间为 0.51s，依然优于所有对比方法。

表 4 运行效率比较（单位：s）。其中，"Ours"表示本文深度神经网络模型，而"Ours+"表示运用额外的 DenseCRF 进一步优化模型输出。

| DRFI | RBD | DS | ELD | DCL | Ours | Ours+ |
|---|---|---|---|---|---|---|
| 6.34 | 0.52 | 0.73 | 0.59 | 1.17 | 0.02 | 0.51 |
| CPU | CPU | GPU | GPU | GPU | GPU | GPU |

**收稿日期**：2018-01-19；**修回日期**：2018-03-15　**基金项目**：广东省应用型科技研发专项资金资助项目（2015B010131002）；广东省科技计划资助项目（2016A040403004）；东莞市重大科技项目（2015215102）

**作者简介**：李鑫（1986-），男，四川江油人，博士研究生，主要研究方向为计算机视觉（xinli_uestc@hotmail.com）；陈雷霆（1966-），男，重庆人，教授，博士，主要研究方向为多媒体应用；蔡洪斌（1966-），男，四川广安人，教授，博士，主要研究方向为三维图形.

## 3　结束语

本文提出一种基于深度学习的显著性对象检测算法。不同于现存基于深度学习的显著性对象检测算法，本文算法在模型设计中考虑了对象结构信息的感知，从而克服了检测复杂显著性对象时结构保存不完整的问题，获得了更高的准确率。为了获取对象的结构信息，本文设计了一个多流的深度神经网络，并且在网络中多尺度融合学习的骨骼信息。除此之外，为了进一步提高显著性图的准确性，本文运用全连接条件随机场优化深度模型的检测结果。实验结果表明，比较传统显著性对象检测算法和现有基于深度学习的检测算法，本文算法在定量和定性指标上都有改善，也说明了在显著性对象检测中考虑对象结构信息的必要性。如何把基于对象结构信息的深度网络模型运用在更多的像素级预测任务，如语义分割和场景分割，以及如何更好地在深度模型中消除冗余信息错误!未找到引用源。是未来工作值得关注的方向。

## 参考文献：


[1] Yang Fan, Li Xin, Cheng Hong *et al.* Object-aware dense semantic correspondence [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017.

[2] Chu Xiao, Yang Wei, Ouyang Wanli *et al.* Multi-Context Attention for Human Pose Estimation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017

[3] Borji A, Frintrop S, Sihite D. N, *et al.* Adaptive object tracking by learning background context [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2012: 23-30.

[4] 徐新，穆楠，张晓龙. 一种鲁棒的夜间图像显著性对象检测模型 [J]. 软件学报, 2018, 29 (9): 1-16. (Xu Xin, Mu Nan, Zhang Xiaolong. A robust salient object detection model for nighttime images [J]. Journal of Software, 2018, 29 (9): 1-16)

[5] 周洋，何永健，刘晓琪，唐向宏，殷海兵. 结合立体视觉舒适度的立体图像显著性检测 [J]. 软件学报, 2017, 28 (s2): 1-10. (Zhou Yang, He Yongjian, Liu Xiaoqi *et al.* Saliency detection for stereoscopic images by considering stereo visual comfort [J]. Journal of Software, 2017, 28 (s2): 1-10)

[6] Geng Yujie. leveraging stereopsis for saliency analysis [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2012: 454–461.

[7] Kim J and Grauman K. Shape sharing for object segmentation [C]// Proc of European Conference on Computer Vision. [S. l. ] : Springer, 2012: 444–458.

[8] Klein D A, Frintrop S. Center-surround divergence of feature statistics for salient object detection [C]// Proc of International Conference on Computer Vision, NJ: IEEE, 2011: 2214–2219.

[9] Li Xin, Yang Fan, Chen Leiting *et al.* Saliency transfer: an example-based method for salient object detection [C]// Proc of International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2016: 3411-3417.

[10] Li Guanbin and Yu Yizhou. Visual saliency based on multiscale deep features [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015

[11] Wang Lijun, Lu Huchuan, Xiang Ruan *et al.* Deep networks for saliency detection via local estimation and global search [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015

[12] Zhao Rui, Ouyang Wanli, Li Hongsheng *et al.* Saliency detection by multi-context deep learning [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2015

[13] Wang Linzhao, Wang Lijun, Lu Huchuan *et al.* Saliency detection with recurrent fully convolutional networks [C]// Proc of European Conference on Computer Vision. [S. l. ] : Springer, 2016: 825–841.

[14] Hou Qibin, Cheng Ming Ming, Hu Xiaowei *et al.* Deeply Supervised Salient Object Detection With Short Connections [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017

[15] Yang Fan, Li Xin, Cheng Hong *et al.* Multi-scale cascade network for salient object detection [C]// Proc of ACM International Conference on Multimedia. New York: ACM Press, 2017.

[16] Li Xi, Zhao Liming, Wei Lina *et al.* DeepSaliency: Multi-task deep neural network model for salient object detection. [J]. IEEE Trans on image processing, 2016, 25 (8): 3919–3930.

[17] Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition [C]// Proc of International Conference on Learning Representations. 2015

[18] Xie Yulin, Lu Huchuan and Yang Minghsuan. Bayesian saliency via low and mid level cues [J]. IEEE Trans on Image Processing. 2013, 22 (5): 1689-98.

[19] Yang Chuan, Zhang Lihe, Lu Huchuan *et al.* Saliency detection via graph-based manifold ranking [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2013: 3166-3173.

[20] Zhu Wangjiang, Liang Shuang, Wei Yichen *et al.* Saliency optimization from robust background detection [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2014: 2814-2821.

[21] Jiang Huaizu, Wang Jingdong, Yuan Zejian *et al.* Salient object detection: a discriminative regional feature integration approach [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2013: 2083-2090.

[22] Lee G, Tai Y W, Kim J. Deep saliency with encoded low level distance map and high level features [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016: 660-668.

[23] Li Guanbin, Yu Yizhou. Deep contrast learning for salient object detection

[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2016.

[24] Dai Jianhua, Wei Bingjie, Shi Hong *et al.* Uncertainty measurement for incomplete interval-valued information systems based on α-weak similarity [J]. Knowledge-Based Systems, 2017, 136: 159-171.